DOCUMENT RESUME

ED 113 395 TM 004 899

AUTHOR Anderson, Edwin R.
TITLE Personal Inquiry in the Classroom: An Alternative Approach to Educational Research. Report No. 76-5.
INSTITUTION Washington Univ., Seattle. Educational Assessment Center.
REPORT NO EAC-76-5; EAC-P-4
PUB DATE Aug 75
NOTE 12p.

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS *Academic Achievement; Analysis of Variance; *Behavioral Objectives; *Classroom Research; College Students; College Teachers; Higher Education; Instructional Design; Research Methodology; *Student Testing; *Teacher Role

ABSTRACT

A method of individual instructor classroom research is proposed and demonstrated in this paper. By capitalizing on baseline data collected from a similar, prior class and focusing treatment on difficult test items, the instructor can develop sensitive within class designs for the detection of treatment effects. An experiment using written objectives given to the students to aid their learning and using the single classroom design shows that objectives significantly improve learning. Written objectives are also shown to interact with item difficulty. The experiment leads to the recommendation that instructors engage in personal inquiry into their efforts to influence student learning and that such inquiry be added to the educational research literature. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED113395

Educational Assessment Center

University of Washington

August 1975

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

76 - 5

Personal Inquiry in the Classroom:

An Alternative Approach to Educational Research

Edwin R. Anderson

Abstract

A method of individual instructor classroom research is proposed and demonstrated in this paper. By capitalizing on baseline data collected from a similar, prior class and focussing treatment upon difficult test items, the instructor can develop sensitive within class designs for the detection of treatment effects. An experiment using written objectives given to the students to aid their learning and using the single classroom design shows that objectives significantly improve learning. Written objectives are also shown to interact with item difficulty. The experiment leads to the recommendation that instructors engage in personal inquiry into their efforts to influence student learning and that such inquiry be added to the educational research literature.

Educational Assessment Center Project: 4

TM 004 899

Personal Inquiry in the Classroom: An Alternative Approach to Educational Research

Inferential statistics has been developed to solve, in a quantitative manner, the problems inherent in generalizing from a specific sample to an abstract population. Even with the cautions provided by inferential statistics, a common error made in their use is overgeneralization. This error suggests that we hold the goal or value of explaining the world in general too dearly and consequently bias our thinking in the direction of too much abstraction. In educational research terms this translates into trying to find instructional methods appropriate in all classrooms for all instructors. Perhaps this dream is too grandiose. The individual teacher is primarily concerned with his class, his students, and his ability to influence those students in ways he judges to be favorable. The teacher is engaged in the exercise of personal influence and needs a method of inquiry that will allow him to assess the outcome of his particular influence attempts. The only generalization he needs to be concerned with is that of transferring his impact to a new group of students.

Educational research needs to move to the individual instructor's classroom if rapid progress on theories of instruction is to be made possible. There are two reasons for such a move. (a) The cost of doing research with many classrooms, several teachers, and large numbers of students is too great; we will never have a large number of such studies. This approach to research is surely needed; but because of its limits, effort and opportunity, cannot be the major source of educational data. (b) The short laboratory experiment on the other hand should not serve as the major data base for instructional theory because of the possible distortions introduced into the data due to the small size and simplicity of the laboratory situation. For example, a subject asked to learn short passages of prose in a one hour period followed by immediate testing may not engage in the same behaviors he would use in reading a textbook over a period of several weeks. While the laboratory can certainly suggest potentially relevant variables to the individual instructor, the expectation of generalization to the classroom, while positive, must remain low until extensive classroom tests have been made. However, the individual classroom has the right size and complexity from the viewpoint of the learning process and is the right size from the viewpoint of convenience in the research effort.

If the individual instructor is to be the principal researcher, he needs a sharper set of research methods than is normally provided. The stability of item difficulties from quarter to quarter and class to class opens the possibility of improved research techniques (Anderson, 1975). In the Anderson study, item difficulty is defined as the proportion of students choosing the correct response to a test item. Item difficulties were calculated for identical examinations given to students in several sections of a Fortran programming class and these item difficulties were then correlated on an item by item basis. The average intercorrelation of five classes in the autumn of 1974 was .73 and the average intercorrelation of eight classes in the winter of 1975 was .71. These high correlations were found in spite of considerable variation in the sections, e.g., different instructors, different practice problems, different lectures, etc. These findings indicate that material easily learned by one class will be easily learned by another class and that difficult material in one class will prove difficult in a second class. Because of the commonality demonstrated by the high correlations between classes, we would expect data from one quarter to be useful as a reliable baseline for data gathered in subsequent quarters. The high correlations encourage the subtraction of baseline item difficulties from item difficulties obtained during the treatment quarter in order to arrive at a change score which more sensitively reflects the effect of the treatment.

The effect of changing classes (different subjects) can be ruled out by applying the treatment to part of the to-be-learned material. This step allows a within class comparison of the change from baseline for untreated and treated item difficulties.

One final touch is needed because the item difficulties for a test given in a live classroom tend to form a skewed distribution with the peak above the mean difficulty. Items which baseline at high difficulty values (easy items) do not have much room for change, i.e. there may be a ceiling effect on the majority of the test items. The group of items selected for treatment should thus consist of one-half of the items having low difficulties in the baseline period. Treatment is confined to one-half of the difficult items because the remaining difficult items are needed as a control for regression effects. Baselining item difficulties, within class design, and focus of treatment analysis on difficult items should all contribute to the sensitivity of the

classroom design as a research tool. (These design features strongly resemble the research methods developed by the behavior modifiers, Bandura, 1969, and their work should provide many useful cues to the classroom instructor engaged in inquiry into his personal effectiveness.)

The remainder of this paper illustrates the classroom design features just mentioned in connection with a specific treatment. Duchastel and Merrill (1973) reviewed several classroom and laboratory studies which assessed the effectiveness of written objectives in improving student performance. Approximately 45% of these studies showed objectives to be beneficial. However, many of the comparisons made in the classroom were poorly controlled, i.e., different classes were used for control and treatment, item difficulties were not used, and there was no focus on difficult items. The purpose of this study is thus two fold: (a) to provide a demonstration of an individual classroom research methodology and (b) to determine whether or not presenting students with written objectives will favorably influence their learning.

Methods

Subjects. Baseline item difficulties were collected from 64 Introductory Psychology students at Shoreline Community College in Seattle, Washington. There were two separate but similar sections of the introductory class; each section was given the same tests and the data from the two sections was pooled for scoring. The experiment proper was conducted with 54 students in an evening Introductory Psychology class at the University of Washington.

Course Materials. The reading assignments for both courses were taken from Beach, Psychology: Core Concepts and Special Topics, 1973, and from a book of readings provided with the test (Sjursen & Beach, Readings in Psychology: Core Concepts and Special Topics, 1973). The page numbers of the reading assignments were given to the students at least two weeks prior to the testing dates. The rest of the content of the course was presented via lectures. An effort was made to prevent overlap in the content of the readings and lectures and the author estimates that the effort was 85% successful.

Research design. During the first four weeks of the experimental quarter (U of W), the students were given, at the beginning of each lecture, a mimeographed sheet which had statements of the major points of the lecture written as objectives. For example,

Each student will recognize an unobtrusive measure as a measurement which does not alter real world events during its use in measuring.

Given a new experiment testing Barker's ecological theory and the knowledge that the experiment confirmed the theory, each student will recognize the outcome of the experiment.

The students were given 63 such objectives during the four week period. No objectives were provided for the reading assignments during this period. In the second four weeks of the quarter a reversal occurred; 50 objectives were given with the reading assignments and no objectives were given with lectures. There was no experimentation or data analysis during the last two weeks of the quarter or during the week of final examinations. The objectives were written after the text and lectures were written and so they served primarily to make explicit the goals for learning implicitly embodied in the written material.

Student performance was measured with multiple choice test questions. The questions which corresponded to the example objectives are as follows:

A museum director measures the wear and tear on the floor tiles in front of pictures to assess their popularity. This is an example of...

A. an unobtrusive measure.
B. operator generated data.
C. a discriminative stimuli.
D. an experimental manipulation.

Wicker tested Barker's conclusions concerning the undermanning and overmanning of behavior settings in small (338 member) and large (1559 member) churches. The churches were both in the same large city, were the same denomination, and had members from the same socioeconomic class. His findings agreed with Barker's population studies. Which of the following is true of Wicker's findings?

A. The large church offered a greater variety of behavior setting.
B. The members of the small church participated in a greater variety of behavior settings.
C. Small church members donated more money per year.
D. All of the above.

The responses of the experimental University of Washington students were turned into an item difficulty (proportion of students responding correctly) from which the item difficulty for the Shoreline students was subtracted. The resulting

change scores were separated on two dimensions, (a) presence or absence of objectives and (b) high (above 60%) or low (below 59%) item difficulty, and analyzed with analysis of variance techniques. The high or low difficulty split was done using the baseline difficulties collected at Shoreline Community College.

Results

There were 37 test items in common between the Shoreline and University of Washington classes during the first four week period, 22 of those related to the text and 15 related to the lectures. The 22 items related to the text had a Pearson product moment correlation of .83 between the two classes whereas the 15 items related to the lectures (objectives present) correlated .52. For the second four week period, there were 40 common items, 22 related to the text (objectives present) and 18 to the lectures. In this period the lecture question item difficulties correlated .79 and the text question item difficulties correlated .52. Note that where no changes were introduced to the course content the correlation of the Shoreline baseline difficulties with the University of Washington difficulties is high (.83 and .79) whereas change efforts lowered the correlations obtained (.52 and .52). Considerable commonality exists in the performance of these two classes.

More to the point are the results of a two X two unweighted means analysis of variance (Myers, 1972, p. 116) performed on the difference scores. Recall that each difference score was obtained by subtracting the item difficulty of the i^{th} item from the Shoreline class from the item difficulty of i^{th} item at the University of Washington. Table 1 shows the mean change scores for each cell while table 2 reports the analysis of variance data. The presence of objectives significantly improves student performance, the change scores are clearly larger with more difficult items, and the significant objectives X difficulty interaction shows that the objectives had their biggest impact on the difficult items. These effects are discoverable with the design used even though the University of Washington students performed in general at a higher level.

Table 1

Change Score Means

	Proportion Correct at Shoreline	
	0 - .59	.60 - 1.00
Objectives present	.29 *N = 19	.7 N = 18
No objectives	.12 N = 14	.5 N = 26

*N is the number of test question change scores included in the cell.

Table 2

Analysis of Variance Data

SV	df	SS	MS	F
Objectives	1	.009	.009	11.39**
Difficulty	1	.021	.021	26.58**
Obj X Diff	1	.0056	.0056	7.09*
S/Obj X Diff(Adj)	73	.0578	.00079	

*$p < .01$

**$p < .001$

Discussion

The methods of experimental analysis used in this study showed the presence of objectives to make a significant difference in the performance of introductory psychology students ($p < .001$). It is interesting to note that without the sensitivity of the analysis performed these differences would not have been found. An independent t test comparing the item difficulties of items with objectives with the item difficulties of the items without objectives showed no significant difference between the two types of items ($t_{75} = 1.02$, $p > .10$). The finer analysis allowed by using baseline data from previous classes is clearly needed if we are to assess the effect of single treatment variables in complex classroom situations.

The interaction of item difficulty with the written-objectives treatment points to a common weakness in the design of classroom experiments. We need to focus our analysis on the more sensitive test items if we are to detect the full effect of our treatment. Ceiling effects work against the detection of differences between the treatment and a control condition and lead us to conclude that our treatment is not effective when in fact it may be quite effective. This interaction clearly points to the need for instructional improvement efforts which focus on selected segments of course content, and so we are led to the same focus on difficult material in the classroom that paired-associate

researchers (Atkinson, 1972; Atkinson & Paulson, 1972) have used in laboratory studies.

There are limitations to this type of study. Will the results generalize? It is possible that the author has a talent for writing objectives which other instructor's do not share, but that does not seem likely. The objectives are nothing more than a statement of the key concepts the students are to learn. Perhaps the effect of objectives discovered here is partly due to the nature of introductory psychology or the multiple choice testing format. Until this experiment is conducted in other classrooms, with different content and different testing methods, the question, "Will the results generalize?", cannot be rigorously answered. Certainly the results of this one experiment should lead to the positive expectation that objectives (or any other variable producing positive results in the classroom) will make a difference elsewhere. A second problem concerns the inability of the single classroom design to introduce variability on dimensions which may be relevant to student performance. For example, the individual instructor would normally be seen as having a single, unvariable constellation of personality traits. Research with variables like personality will have to involve more than one instructor.

All activities engaged in by humans are at some fundamental level personal. The individual instructor teaches his students, his course and this personal environment is where the instructor must strive to improve his ability to teach. If he wants to be systematic about such improvement efforts, he needs research designs which allow personal inquiry. Generalization to other instructor's classes, while it may occur, is not the issue. In making the individual classroom the location of emphasis in the educational research effort, we are focussing on the one place where research can find a vigorous home. The number of opportunities for such research is large and the effort needed is often no more than the effort teachers normally expend to improve their courses. Eventually enough single case experiments such as the experiment reported here will accumulate so that generalizations may appear. Even if the research outcomes prove situation specific, the methods used to discover the particular results will generalize. The instructor need only adapt the research methods to his personal inquiry.

References

Anderson, E. R. The certainty of information in instructional decision making. University of Washington Educational Assessment Center project report 503, August 1975.

Atkinson, R. C. Optimizing the learning of a second language vocabulary. Journal of Experimental Psychology, 1972, 96, 124-129.

Atkinson, R. C. & Paulson, J. S. An approach to the psychology of instruction. Psychological Bulletin, 1972, 78, 49-61.

Bandura, A. The Principles of Behavior Modification. New York: Holt, Rinehart, & Winston, 1969.

Beach, L. R. Psychology: Core Concepts and Special Topics. New York: Holt, Rinehart, & Winstron, 1973.

Duchastel, P. C. & Merrill, P. F. The effects of behavioral objectives on learning: a review of empirical studies. Review of Educational Research, vol. 43, 1973, 53-69.

Myers, J. L. Fundamentals of Experimental Design. Boston: Allyn and Bacon, 1966.

Sjursen, F. & Beach, L. R. Readings in Psychology: Core Concepts and Special Topics. New York: Holt, Rinehart, & Winston, 1973.